REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)* 14-09-2015	**2. REPORT TYPE** Final	**3. DATES COVERED** *(From - To)* 22 May 2014 to 21 May 2015

4. TITLE AND SUBTITLE

Secure Data Aggregation Protocol for M2M Communications

5a. CONTRACT NUMBER
FA2386-14-1-4029

5b. GRANT NUMBER
Grant 14IOA066_114029

5c. PROGRAM ELEMENT NUMBER
61102F

6. AUTHOR(S)

Prof. Rongxing Lu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Nanyang Technological University
50 Nanyang Avenue
Singapore 639798
Singapore

8. PERFORMING ORGANIZATION REPORT NUMBER

N/A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

AOARD
UNIT 45002
APO AP 96338-5002

10. SPONSOR/MONITOR'S ACRONYM(S)

AFRL/AFOSR/IOA(AOARD)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
14IOA066_144029

12. DISTRIBUTION/AVAILABILITY STATEMENT

Distribution Code A: Approved for public release, distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Data aggregation is an important data process technique in Machine to Machine (M2M) communications, which, upon the request from some specific application requirement, only transmits the selected/processed data to the application domain.

15. SUBJECT TERMS

Aggregation Protocol, Data Integrity, Intelligent Device, Fault-Tolerant Computing, Multi-dimensional Data

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Seng Hong, Ph.D.
U	U	U	SAR	25	**19b. TELEPHONE NUMBER** *(Include area code)* +81-4-2511-2000

Report Documentation Page	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE **14 SEP 2015**	2. REPORT TYPE **Final**	3. DATES COVERED **22-05-2014 to 21-05-2015**

4. TITLE AND SUBTITLE **Secure Data Aggregation Protocol for M2M Communications**	5a. CONTRACT NUMBER **FA2386-14-1-4029**
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER **61102F**
6. AUTHOR(S) **Rongxing Lu**	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Nanyang Technological University,50 Nanyang Avenue,Singapore 639798,Singapore,NA,NA**	8. PERFORMING ORGANIZATION REPORT NUMBER **N/A**
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) **AOARD, UNIT 45002, APO, AP, 96338-5002**	10. SPONSOR/MONITOR’S ACRONYM(S) **AFRL/AFOSR/IOA(AOARD)**
	11. SPONSOR/MONITOR’S REPORT NUMBER(S) **14IOA066_144029**

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Data aggregation is an important data process technique in Machine to Machine (M2M) communications, which upon the request from some specific application requirement, only transmits the selected/processed data to the application domain.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a REPORT **unclassified**	b ABSTRACT **unclassified**	c THIS PAGE **unclassified**	**Same as Report (SAR)**	**25**	

Final Report for AOARD Grant FA2386-14-4029

"Secure Data Aggregation Protocol for M2M Communications"

Date March 24th, 2015

Name of Principal Investigators (PI and Co-PIs): Rongxing Lu
- E-mail address: rxlu@ntu.edu.sg
- Institution: School of Electrical and Electronics Engineering, Nanyang Technological University
- Mailing Address: 50 Nanyang Avenue, Singapore 639798
- Phone: +65 6790-4519
- Website: http://www.ntu.edu.sg/home/rxlu/

Period of Performance: May/22/2014 – May/21/2015

Abstract:

Data aggregation is an important data process technique in M2M communications, which, upon the request from some specific application requirement, only transmits the selected/processed data to the application domain. In this research proposal, we aim to investigate secure data aggregation with fault tolerance in M2M communications. Different from previous research works on secure data aggregation, information privacy and data integrity will be simultaneously integrated in data aggregation, and fault tolerance will be also studied in this proposal. In specific, the novelty of this research project lies in the following aspects: i) develop new data aggregation schemes to simultaneously achieve the information privacy and data integrity in M2M communications; ii) develop new privacy-preserving data aggregation schemes with fault tolerance for M2M communications. This proposal will contribute to the secure communications and information exchange between sets of nodes, and the lessons learned will also better prepare AOARD for establishing the strategy towards new information security and transmission challenges in future military M2M communications.

Introduction:

M2M communication is characterized by involving a large number of intelligent devices sharing information and making collaborative decisions without direct human intervention. Due to its potential to support a large number of ubiquitous characteristics and achieving better cost efficiency, M2M communication has quickly become a market-changing force for a wide variety of real-time monitoring applications, such as traffic surveillance, smart metering, environmental monitoring, industrial automation and military scenarios [1][2]. Despite various M2M applications, the basic M2M communication infrastructure is quite similar and usually consists of three parts: M2M domain, network domain, and application domain [3], as shown in Fig. 1. In the infrastructure, the information is generated by sensors in M2M domain, then transmitted through wire/wireless network in network domain, and finally through a gateway to application domain, where it can be reviewed and acted on. To support this kind of information flow in M2M communication, data transmission is a critical component in the infrastructure. However, due to huge data generated at M2M domain, it is infeasible or cost-inefficient to directly transmit these high-volume data. Therefore, transmission of selected/processed data is expected in M2M communications.

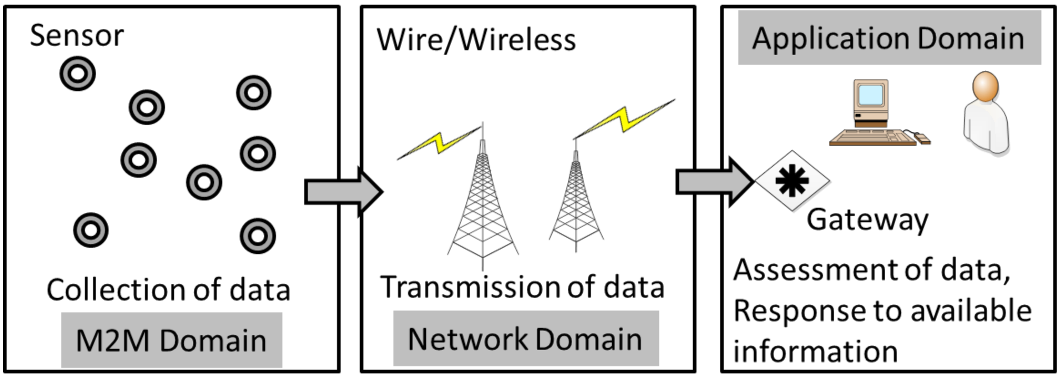


Fig. 1. Basic M2M communication infrastructure

Data aggregation [4] is an important data process technique, which, upon the request from some specific application requirement, only transmits the selected/processed data, e.g., **count, sum, max, min, average** values, to the application domain. Therefore, it can largely reduce the transmission cost while still meeting the application requirement. Over the past years, due to its efficiency, data aggregation has been paid great attention, and plentiful data aggregation schemes have been proposed [4][5]. However, many previously reported data aggregation schemes cannot be directly applied to M2M communications, partly because they did not take the unique characteristics of M2M communications into good consideration. Since sensors in M2M communications are usually low cost, small size and often deployed at unattended environments, they are easily vulnerable to malicious attacks and/or sometimes malfunctioning [3]. Therefore, in order to make the data aggregation really workable in M2M communications, the requirements of security and fault tolerance should be reinforced in data aggregation. We note that, although some secure data aggregation schemes [6][7][8] were proposed in sensor networks to resist pollution attacks, they sometimes do not work well due to the lack of the fault tolerance property. Therefore, secure data aggregation with fault tolerance still needs further study in M2M communications.

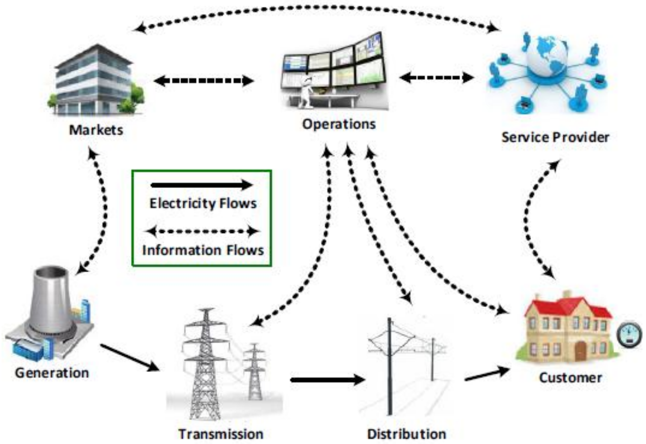
[1] M2M Communications, http://www.m2mcomm.com/
[2] How Machine-to-Machine Communication Works, http://computer.howstuffworks.com/m2m-communication1.htm
[3] R. Lu, X. Li, X. Liang, X. Lin, and X. Shen, "GRS: The Green, Reliability, and Security of Emerging Machine to Machine Communications", IEEE Communications, Vol. 49, Issue 4, pp. 28-35, 2011.
[4] R. Rajagopalan and P. K. Varshney, "Data aggregation techniques in sensor networks: A survey." (2006).
[5] W. Guo et al. "Multi-source temporal data aggregation in wireless sensor networks." Wireless personal communications 56.3 (2011): 359-370.
[6] R. Bista and J.W. Chang, "Privacy-preserving data aggregation protocols for wireless sensor networks: a survey." Sensors 10.5 (2010): 4577-4601.
[7] M. Joye and B. Libert, "A Scalable Scheme for Privacy-Preserving Aggregation of Time-Series Data", in FC 2013.
[8] J. Shi et al. "Prisense: privacy-preserving data aggregation in people-centric urban sensing systems." INFOCOM, 2010 Proceedings IEEE. IEEE, 2010.

Experiment, Results and Discussion:

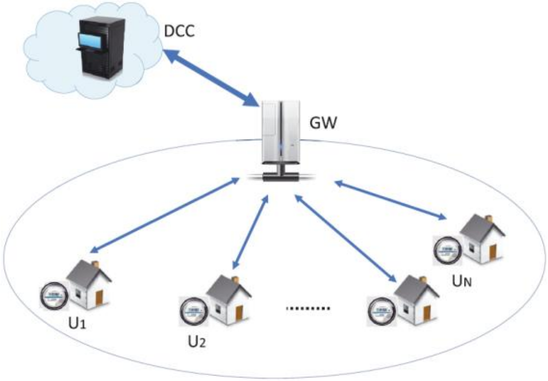
(1) PDA: A Privacy-Preserving Dual-Functional Aggregation Scheme for Smart Grid Communications

Abstract

Privacy-preserving aggregation for smart grid communications, which precisely meets the requirement of periodically collecting users' electricity consumption while preserving privacy of each individual user, has been extensively studied in recent years. However, most of the existing privacy-preserving aggregation schemes only focused on the summation aggregation. In this paper, based on the lattice cryptographic technique, we propose a novel privacy-preserving dual functional aggregation scheme (PDA) for smart grid communications. With our proposed PDA scheme, each individual user just reports one data, then multiple statistic values, i.e., mean and variance, of all users can be computed by the data & control center in the smart grid, while the privacy of each individual user can still be protected. Detailed security analyses demonstrate that our proposed PDA scheme is secure and robust. In addition, extensive performance evaluations also show that our proposed PDA scheme is efficient in terms of computational and communication overhead.



The conceptual smart grid system architecture



System model under consideration

Major Features and Contributions

- PDA uses a homomorphic encryption scheme to encrypt users' data so that the users' privacy can be protected from eavesdropping under the defined attack model. PDA supports both additive and multiplicative aggregations, which enables data & control center (DCC) to obtain both mean and variance of the users' data with only one report sent by each user.
- Additional techniques, including multi-bits ring LWE encryption, encoding integers to polynomials, and super-increasing sequence filling, are integrated into the optimized

PDA to further reduce the computational cost, take full advantage of bandwidth, and ease communication overhead.

Performance Evaluation

Our proposed PDA scheme outperforms the basic Paillier-based aggregation scheme [1,2,4] in terms of computation cost and communication overhead.

A. Computation Cost

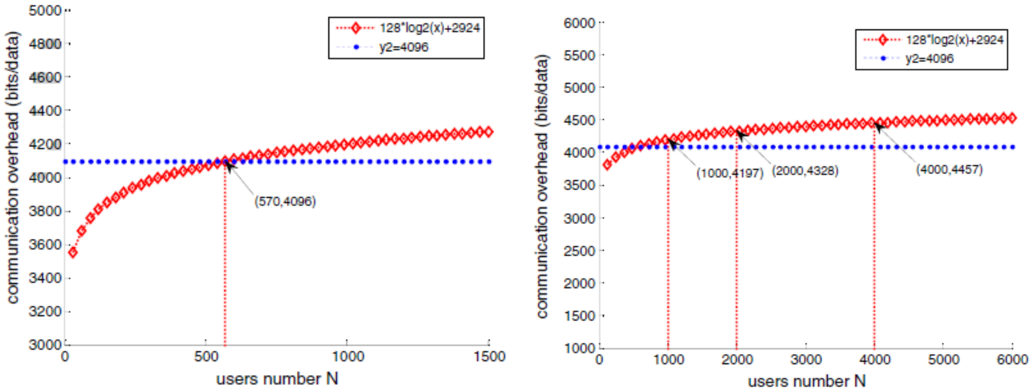
Encryption cost	Paillier encryption scheme	$O(n^2 \log n \log\log n)$
	Our scheme	$O(n \log^2 n)$
Homomorphic operation cost	Paillier encryption scheme	$O(n \log n \log\log n)$
	Our scheme	$O(n \log n)$
Decryption cost	Paillier encryption scheme	$O(n^2 \log n)$
	Our scheme	$O(n \log^2 n)$

B. Communication Cost

Optional parameters setting of our scheme are as follows,

$n = k$	L	$n' = (2^l - 1) * 20$	$n \ \log q/L$
192	3	140	$96 \log N + 1454$
256	3	140	$128 \log N + 2924$
320	4	300	$150 \log N + 3288$

By setting n = 256 for our scheme, it can provide security which is equivalent to AES-128 and is enough for ordinary use. While for Paillier-based scheme, to provide the same security, the RSA modulus N should be set to be 2048 bits. Under such parameters settings, the comparison of communication overhead between our scheme and the basic Paillier-based scheme is shown as follows.

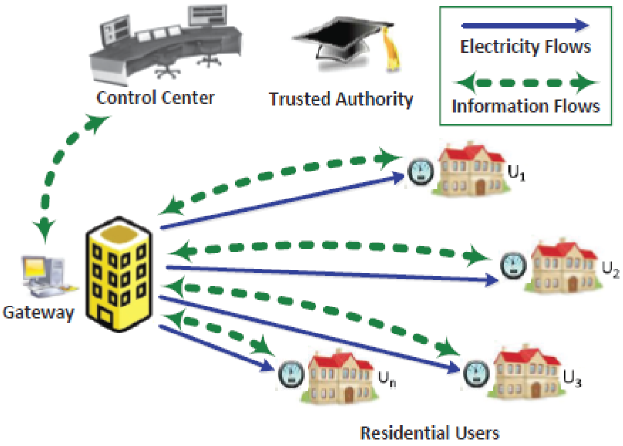


Communication overhead comparison (bits/data)

(2) A New Differentially Private Data Aggregation with Fault Tolerance for Smart Grid Communications

Abstract

Privacy-preserving data aggregation has been widely studied to meet the requirement of timely monitoring measurements of users while protecting individual's privacy in smart grid communications. In this paper, a new secure data aggregation scheme, named DPAFT, is proposed which achieves differential privacy and fault tolerance simultaneously. Specifically, inspired by the idea of Diffie-Hellman key exchange protocol, an artful constraint relation is constructed which is different from all the existing similar works. Thanks to this novel constraint, DPAFT can support fault tolerance of malfunctioning smart meters efficiently and flexibly. DPAFT is also enhanced to resist differential attacks which are suffered by most of the existing data aggregation schemes. Moreover, by improving the basic Boneh-Goh-Nissim cryptosystem to be more applicable to the practical scenarios, DPAFT can resist much stronger adversaries, i.e., the users' privacies are protected in the honest-but-curious model. In addition, extensive performance evaluations are conducted to illustrate that DPAFT outperforms a state-of-the-art data aggregation scheme in terms of storage cost, computation complexity, utility of differential privacy, robustness of fault tolerance, and the efficiency of user addition and removal.



System model under consideration

Major Features and Contributions

- Inspired by the idea of Diffie-Hellman key exchange protocol, we put forward a novel solution for fault tolerance for smart metering. Unlike all of the existing similar works, which depend on the restricted relation of $\sum_{i=0}^{n} s_i = 0$, an artful constraint relation $s_0 \sum_{i=1}^{n} s_i = 1$ is constructed, where s_0, and s_i, for $i = 1,2, \dots, n$, are the private keys of the control center (CC) , and each residential user, respectively.
- By adding Laplacian noise via distributed manner, DPAFT is designed to provide differential privacy by introducing distributed noise generation procedure. Compared with the-state-of-the-art differentially private smart grid aggregation protocol [3], our protocol is more efficient due to the elimination of heavy communication, computation, and storage overhead of future-ciphertexts, while still provides high utility (i.e., low error).
- By improving the basic Boneh-Goh-Nissim cryptosystem to be more applicable to the practical scenarios, our DPAFT can resist much stronger adversary and is highly efficient. Specifically, by hiding the private key p of the basic Boneh-Goh-Nissim cryptosystem to the CC and introducing the blind factor t for the GW and the secret key r for the CC, respectively, the users' electricity usage privacy is protected in honest-but-curious model.

Performance Evaluation

Our proposed scheme is compared with the state-of-the-art scheme proposed by Jongho et al. [3] as follow.

A. Storage Cost

Jongho et al.'s scheme	Huge amount of memory buffers need to be configured for GW to store the future ciphertexts.
Our proposed scheme	GW is just responsible for data aggregation and packages relay. No special storage requirements are needed.

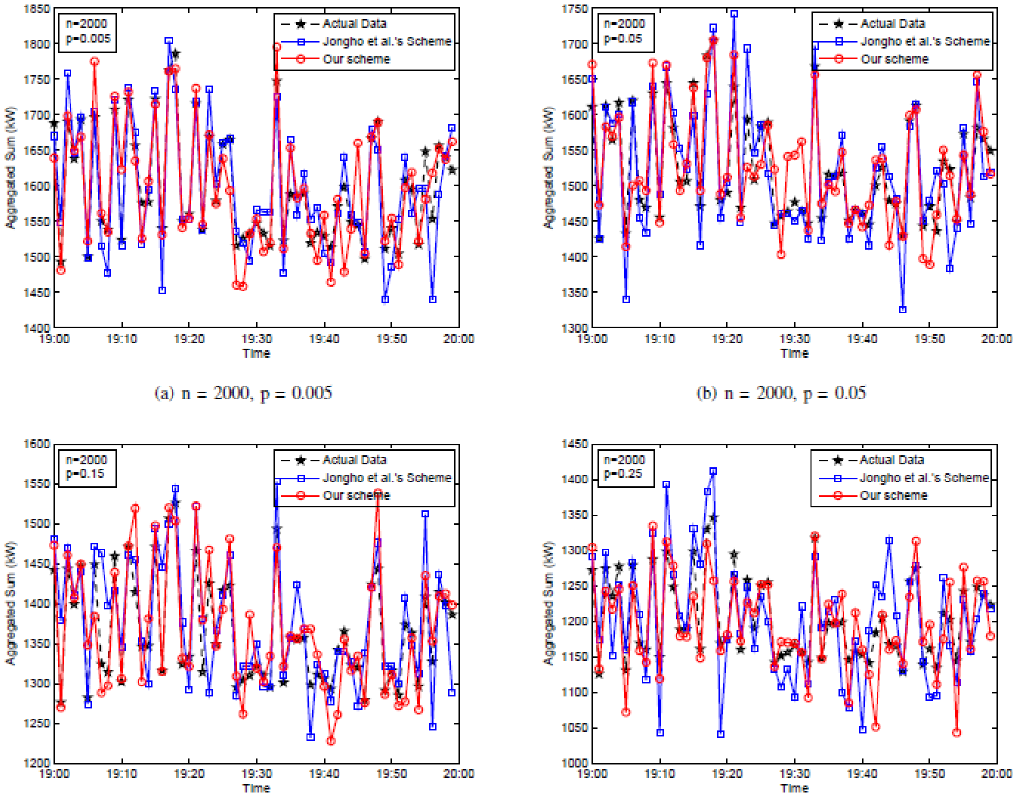
B. Computation Complexity

Jongho et al.’s scheme	The shared secret keys between every two users in the k-size partner group should be generated and assigned secretly. Two parts of ciphertexts, i.e., the current ciphertext and the future ciphertext, should be calculated and reported.
Our proposed scheme	No need to compute and assign the shared secret keys among the users. The additional computation of future ciphertext is not necessary either.

C. Utility of Differential Privacy

Jongho et al.’s scheme	The additional Laplatics noise is added to each smart meter’s future ciphertext to resist the subtracting attack of current ciphertext and future ciphertext, which incurs large errors.
Our proposed scheme	Overcomes the above drawback, thus, it is of better utility.

The following figures compare the actual total measurements, the noisy counterparts of the scheme of Jongho et al. and ours, respectively, where in each of the figure, n and p denote the total number of the household, and the different ratio of malfunctioning smart meters, respectively.



(a) n = 2000, p = 0.005

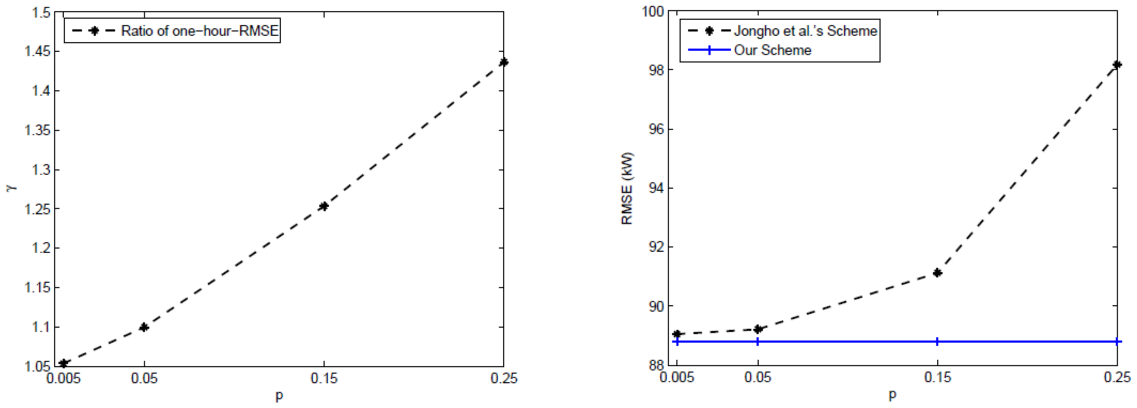
(b) n = 2000, p = 0.05

(c) n = 2000, p = 0.15

(d) n = 2000, p = 0.25

Comparison of noisy total consumption between our scheme and Jongho et al.’s scheme

Define one-day-RMSE (root mean square error), the closeness between the sequences of actual and noisy sums is $\mathrm{RMSE} = \sqrt{\frac{1}{T} \cdot \sum_{t=1}^{T} (\widehat{m_t} - m_t)^2}$, where T = 1440 is the number of time points of one day. Suppose the one-hour-RMSE of Jongho et al.'s protocol and ours be γ_1 and γ_2, respectively. Let $\gamma = \frac{\gamma_1}{\gamma_2}$. The one-hour-RMSE and one-day-RMSE of our scheme and Jongho et al.'s scheme are compared as follows.



Comparison of one-hour-RMSE

Comparison of one-day-RMSE

Comparison of one-hour-RMSE and one-day-RMSE

D. Robustness of Fault Tolerance

Jongho et al.'s scheme	Can only support the maximum $B \cdot T$ long period of fault tolerance, where B is the buffer size of the future ciphertexts for each smart meter and T is the report interval.
Our proposed scheme	Support robust data aggregation with any rational number of malfunctioning smart meters with arbitrary long fault period.

E. Efficiency of User Addition and Removal

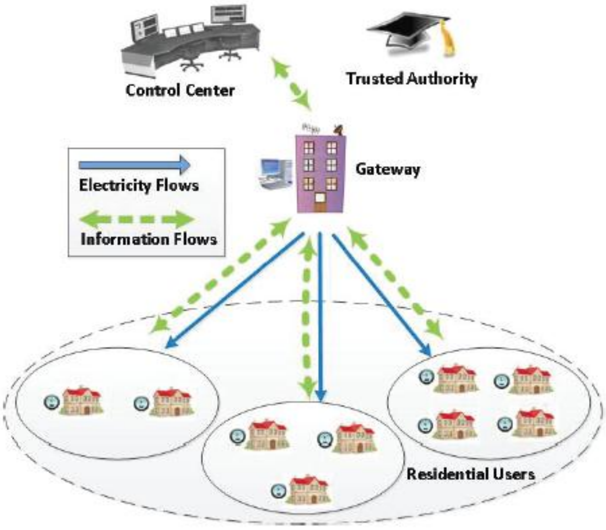
Jongho et al.'s scheme	Costs $O(k \times B)$ and $O(k)$ communication overheads for one user addition and removal, respectively.
Our proposed scheme	Only needs the TA to reassign the key materials for the changed users (user addition and removal).

(3) DDPFT: Secure Data Aggregation Scheme with Differential Privacy and Fault Tolerance

<u>Abstract</u>

A new secure data aggregation scheme, named DDPFT, is proposed for achieving differential privacy and fault tolerance simultaneously. Specifically, by introducing some auxiliary ciphertext subtly, a novel distributed solution for fault tolerant data aggregation is put forward to be able to aggregate the functioning smart meter measurements flexibly and efficiently for any rational number of malfunctioning smart meters with arbitrary long failure period. Furthermore, DDPFT also achieves a good trade-off of accuracy (i.e., low error) and security of differential privacy for arbitrary number of malfunctioning smart meters. Moreover, through decentralizing the computational overhead and the power of the hub-like entity of the gateway, the security of our proposed scheme is enhanced and the efficiency is improved significantly. In addition, extensive performance evaluations are conducted to illustrate that DDPFT outperforms the state-of-the-art data aggregation schemes in terms of computation complexity, communication cost, robustness of fault tolerance, and utility

of differential privacy.



System model under consideration

Major Features and Contributions

- By introducing auxiliary ciphertext subtly, we put forward a novel distributed solution for fault tolerant data aggregation. The fault tolerance mechanism put forward by us is more efficient and robust. By utilizing the auxiliary ciphertexts, the CC can obtain the aggregation of the functioning smart meters flexibly and efficiently for any rational number of malfunctioning smart meters with arbitrary long failure period.
- DDPFT provides differential privacy by adding appropriate noises chosen from symmetric geometric distribution to the aggregation data by the GW. The proposed scheme supports differential privacy and fault tolerance simultaneously and achieves a good trade-off of accuracy and security of differential privacy.
- Through decentralizing the computational overhead and the power of the hub-like entity GW which is usually with limited computation resources and is semi-trust, the security of our proposed scheme is enhanced and the efficiency is improved significantly.

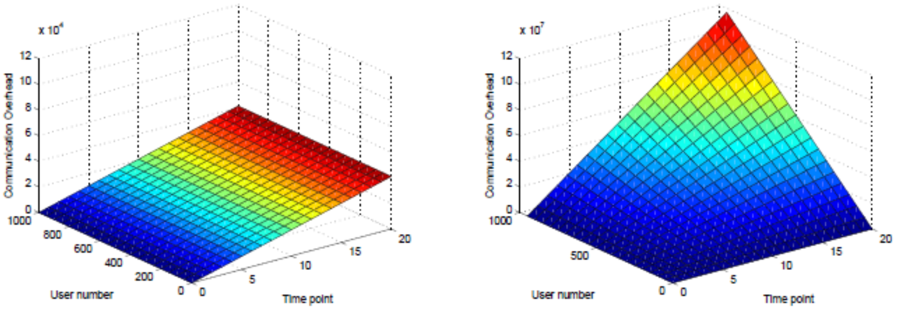
Performance Evaluation

Different from most of the existing similar works, differential privacy and fault tolerance are taken into consideration at the same time in our scheme. We mainly focus on the comparison of our proposed scheme with the state-of-the-art date aggregation schemes supporting differential privacy and/or fault tolerance.

A. Computation Complexity

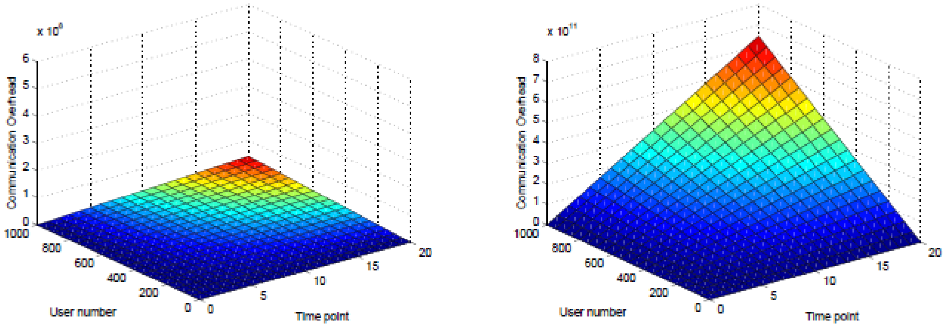
We extend Shi et al.’s privacy-preserving aggregation protocol [4] to support fault tolerance. Our scheme outperforms the scheme of [4] in computation complexity and supports fault tolerance only with a little more computational overhead.

B. Communication Cost



DDPFT Shi et al.'s scheme

Individual communication overhead

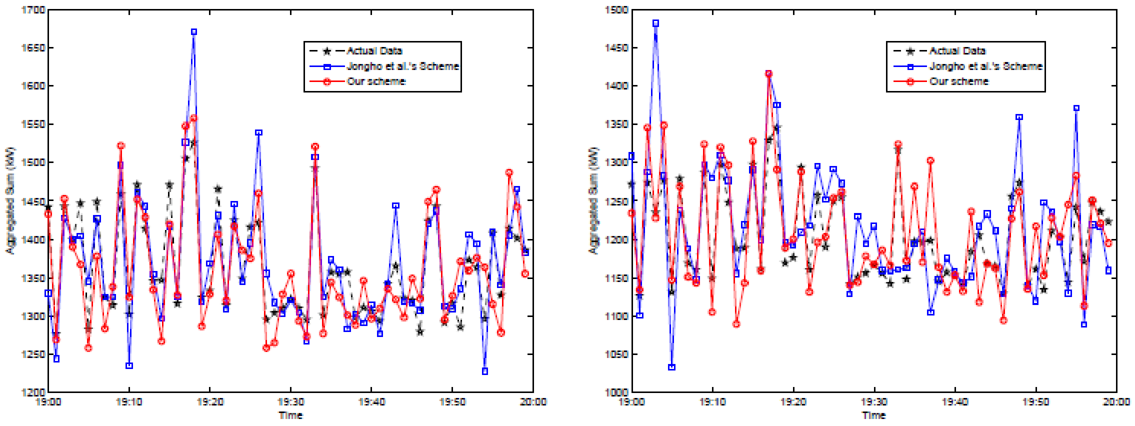


DDPFT Shi et al.'s scheme

Overall communication overhead

C. Utility of Differential Privacy

The proposed scheme provides higher utility (i.e., low error) in terms of differential privacy than the state-of-the-art data aggregation scheme of [3].



(a) $n = 2000,\ p = 0.15$ (b) $n = 2000,\ p = 0.25$

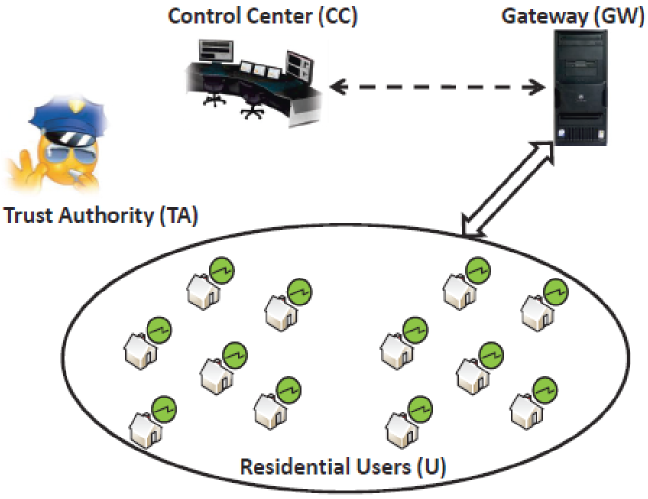
Comparison of noisy total consumption between our scheme and Jongho et al.'s scheme

(4) A Novel Privacy-Preserving Set Aggregation Scheme for Smart Grid Communications

Abstract

In this paper, we propose a novel privacy-preserving set aggregation scheme for smart grid communications. The proposed scheme is characterized by employing a group G of composite order $n = pq$ to achieve two-subset aggregation from a single aggregated data. With the proposed set aggregation scheme, the control center in smart grid is able

to obtain more fine-grained data aggregation results for better monitoring and controlling smart grid. Detailed security analysis shows that the proposed scheme can achieve privacy-preserving property with formal proof in the random oracle model. In addition, extensive experiments are conducted, and the results demonstrate the proposed scheme is also efficient in terms of low computational costs and communication overheads.



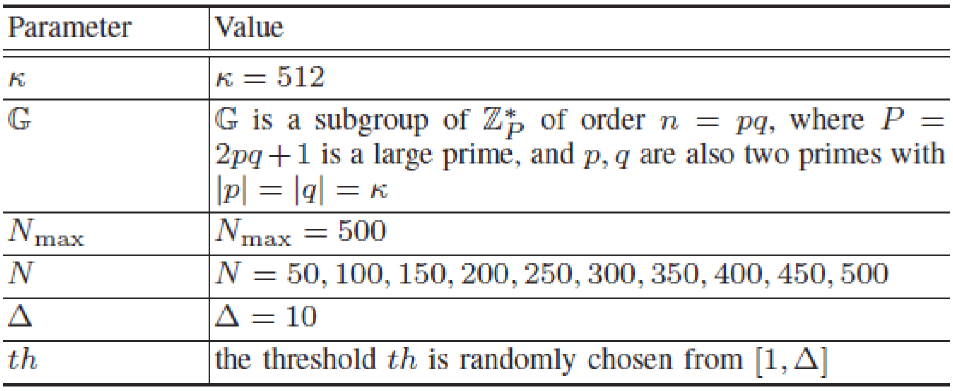
System model under consideration

Major Features and Contributions

- By using a group of composite order, we propose a novel privacy-preserving set aggregation scheme. Given a threshold of electronic consumption data, users can be divided into two subsets, then the proposed scheme can use one single aggregated data to aggregate the sum of electronic consumption data in each subset and the corresponding subset size in a privacy-preserving way, which thus supports more accurate data analytics for controlling and monitoring in smart grid.
- With formal security proof technique, we show our proposed scheme can achieve each individual user's data privacy preservation.
- We implement our proposes scheme in Java and run extensive experiments to validate its efficiency in terms of low computational cost and communication overhead, and discuss the trade-off between the utility and differential privacy level.

Performance Evaluation

We evaluate our proposed privacy-preserving set aggregation scheme in terms of computational cost and communications overheads. Specifically, we implement our scheme by Java (JDK 1.8) and run our experiments on a Laptop with 3.1 GHz processor, 8GB RAM, and Window 7 platform. The detailed parameter settings are shown as follows.

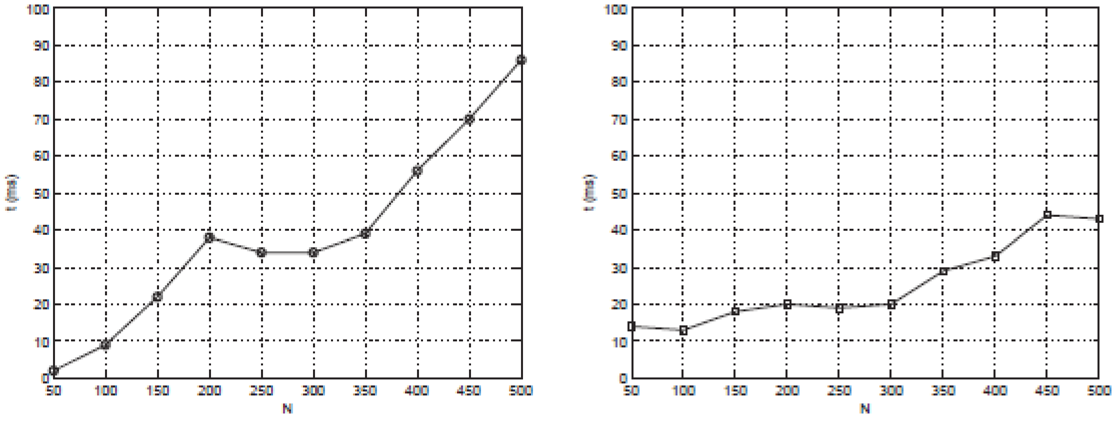


Parameter	Value
κ	$\kappa = 512$
$\mathbb{G}$	$\mathbb{G}$ is a subgroup of $\mathbb{Z}_P^*$ of order $n = pq$, where $P = 2pq+1$ is a large prime, and p, q are also two primes with $\lvert p\rvert = \lvert q\rvert = \kappa$
$N_{\max}$	$N_{\max} = 500$
N	$N = 50, 100, 150, 200, 250, 300, 350, 400, 450, 500$
Δ	$\Delta = 10$
th	the threshold th is randomly chosen from $[1, \Delta]$

Parameter settings

A. Computational Cost

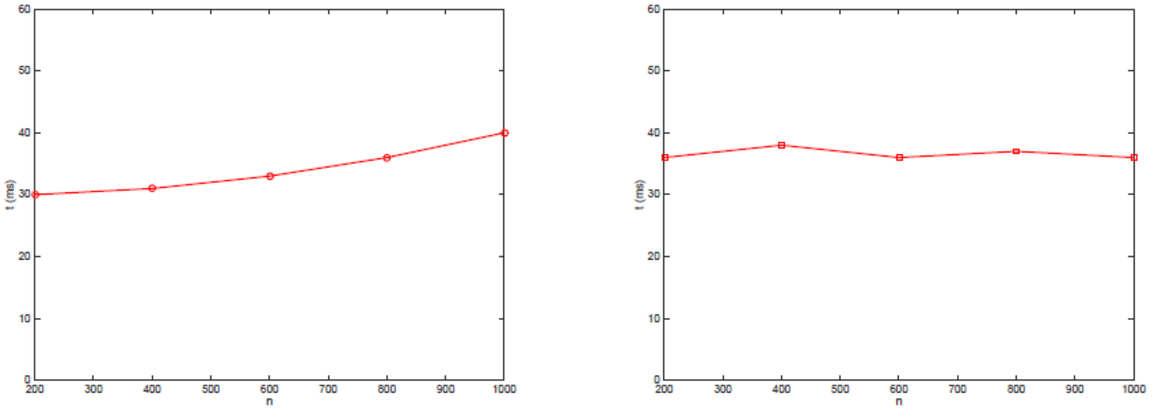
No matter whether a user belongs to subset U_1 or U_0, the average encryption at user side only takes 3.46 ms, which is extremely efficient. The following figure shows the computational costs of aggregation at GW and decryption at CC varies with the number of user N from 50 to 500 with the increment of 50. From the figure, we can see both of them are efficient, and the number of users N has a little effect on the aggregation and decryption, after a hash table used for looking-up in decryption is established in advance.



(a) Aggregation at GW

(b) Decryption at CC

Computational costs of aggregation and decryption varying with N



(a) Aggregation at the gateway

(b) Decryption at the control center

Computational costs of aggregation and decryption varying with n

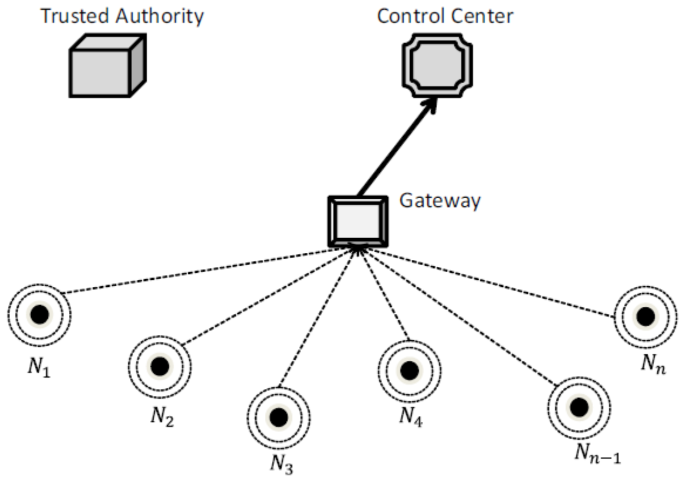
B. Communication Cost

When $|p| = |q| = 512$, the length of $P = 2pq + 1$is 1025 bits. Thus, any ciphertext (including c_i and C) in the subgroup G of Z_P^* is less than or equal to 1025 bits.

(5) Privacy-Preserving Time-Series Data Aggregation for Internet of Things

Abstract

In recent years, the networking and collaboration among various devices has experienced tremendous growth. To adapt to the trend, the concept of Internet of Things (IoT) has been paid great attention not only from the academia but also from the industry. Due to its potential to support a large number of ubiquitous characteristics and achieving better cost efficiency, IoT can find many applications in real world, including traffic surveillance, smart metering, environmental monitoring, industrial automation and military scenarios. Although IoT has attracted a lot of attention; and yet, despite all the attention, has remained many security and privacy challenges. Since most devices in IoT are often deployed at unattended areas, they are vulnerable to the physical attacks while without being detected immediately; and the nature of broadcast in wireless communication also makes an attacker easy to launch eavesdropping attack. As many research efforts have been put on the IoT security challenges, in this chapter, we mainly focus ourselves on addressing the privacy challenges in IoT. To address the privacy challenges, i.e., to protect individual device's data privacy in IoT, many privacy-preserving data aggregation schemes have been proposed. However, most of them only support one-dimensional data aggregation, which sometimes cannot meet the accuracy requirement in IoT scenarios. Although our previous work EPPA deals with the multi-dimensional data aggregation [7], it may not be well support large space data aggregation. Therefore, aiming at the above challenges, we propose a novel privacy-preserving time-series aggregation scheme for IoT, which is characterized by exploiting the properties of group $\mathrm{Z}_{\mathrm{p}^2}^*$ to support data aggregation for both small plaintext space and large plaintext space at the same time, which thus is more efficient than traditional data aggregation.



System model under consideration

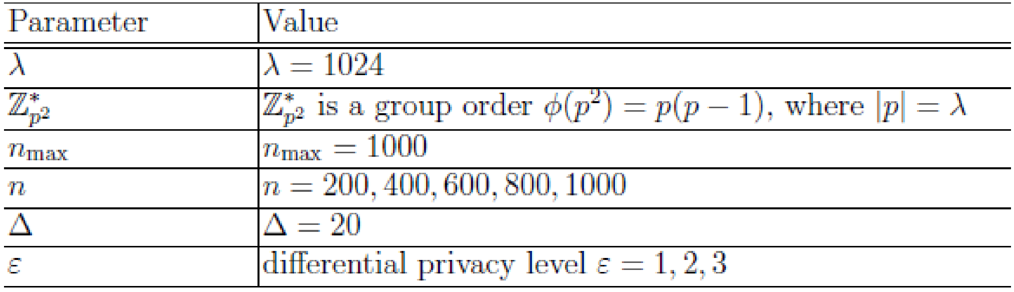
Major Features and Contributions

- We propose a novel privacy-preserving time-series aggregation scheme based on the group $\mathrm{Z}_{\mathrm{p}^2}^*$. The proposed scheme can use one single aggregated data to obtain both the small plaintext space aggregation and the large plaintext space aggregation in a privacy-preserving way at the same time.
- With formal security proof technique, we show our proposed scheme can achieve each individual node's data privacy preservation.
- We implement our proposes scheme in Java and run extensive experiments to validate its efficiency in terms of low computational cost and communication overhead, and discuss the trade-off between the utility and differential privacy level.

Performance Evaluation

We evaluate our proposed privacy-preserving time series aggregation scheme in terms of computational cost and communications overheads, and analyze the utility of differential-privacy as well.

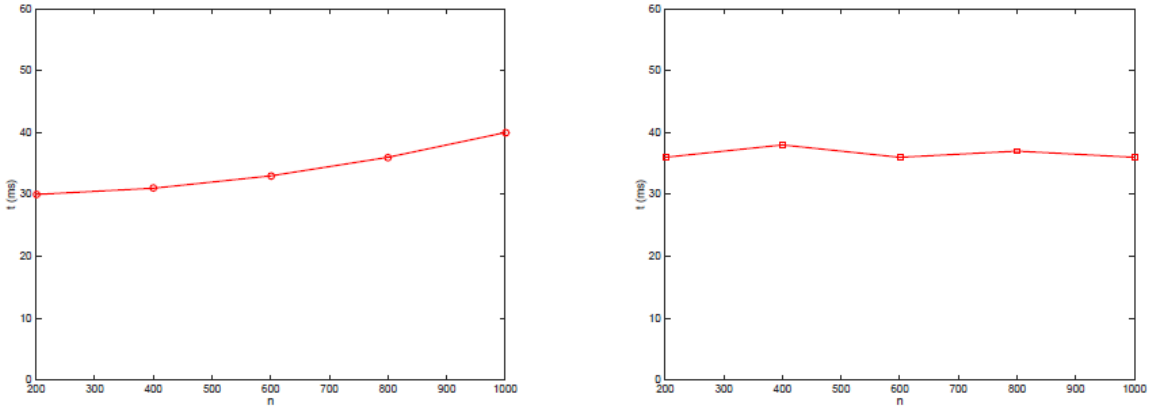
The parameter settings are listed as follows.



Parameter	Value
λ	$\lambda = 1024$
$\mathbb{Z}_{p^2}^*$	$\mathbb{Z}_{p^2}^*$ is a group order $\phi(p^2) = p(p-1)$, where $\|p\| = \lambda$
$n_{\max}$	$n_{\max} = 1000$
n	$n = 200, 400, 600, 800, 1000$
Δ	$\Delta = 20$
ε	differential privacy level $\varepsilon = 1, 2, 3$

Parameter settings

A. Computation Complexity



(a) Aggregation at the gateway

(b) Decryption at the control center

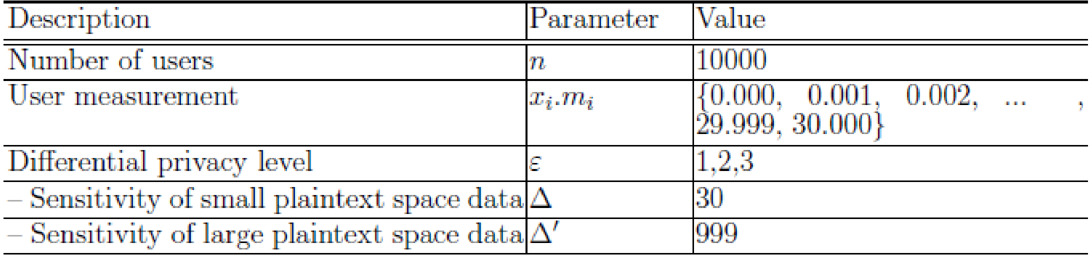
Computational costs of aggregation and decryption varying with n

B. Communication Cost

When $|p| = 1024$, any ciphertext (including c_i and C) in group $Z_{p^2}^*$ is less than or equal to 2048 bits.

C. Utility of Differential Privacy

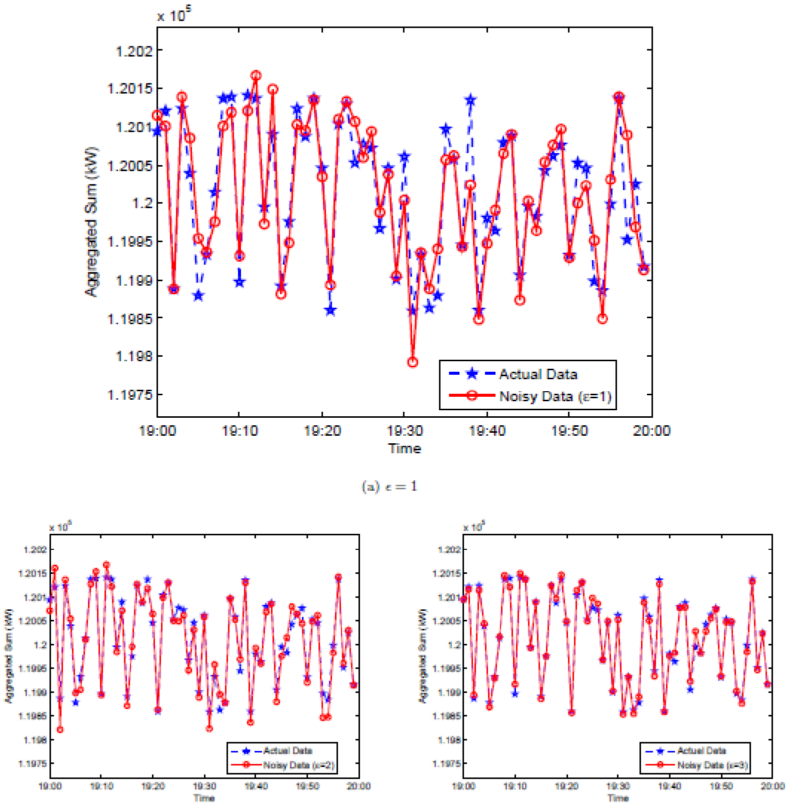
We take smart grid as an example to elaborate the advantages and effectiveness of our proposed scheme. Different from previously reported aggregation schemes for smart grid, our scheme can support data aggregation of user measurements including not only the integer part (small plaintext data $x_i \in [0,30]$) but also the decimal part (large plaintext data $m_i \in [0,999]$). The detailed parameter settings are listed as follows.



Description	Parameter	Value
Number of users	n	10000
User measurement	$x_i.m_i$	{0.000, 0.001, 0.002, ... , 29.999, 30.000}
Differential privacy level	ε	1,2,3
– Sensitivity of small plaintext space data	Δ	30
– Sensitivity of large plaintext space data	Δ'	999

Parameter settings for the evaluation of utility of differential privacy

Based on the real data for 10000 households, we plot the traces of actual total measurements and noisy total consumptions for small plaintext space and large plaintext space, respectively. We also set ε, the differential privacy level, to 1, 2, 3, for each of the two scenarios. As it can be seen from the following figures, the larger ε is, the smaller noise will be added, and then the utility is higher while the smaller ε is, the larger noise will be included, and then the higher level of the privacy can be guaranteed. Compared with the case of ε = 3, the utility in ε = 1 is lower, but it is still acceptable. Therefore, in real scenarios, there is a trade-off between the privacy and utility.

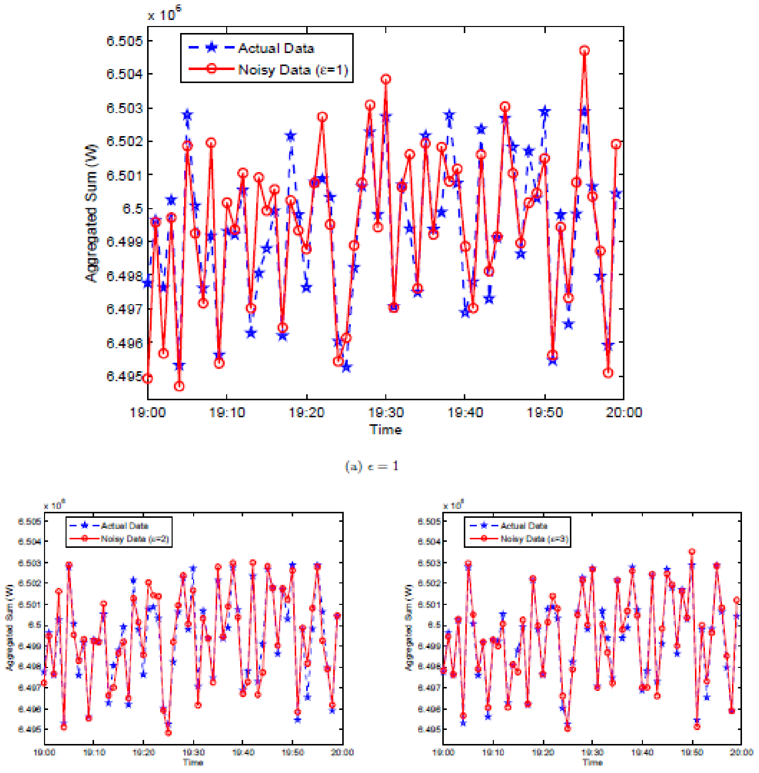


(a) $\epsilon = 1$

(b) $\epsilon = 2$

(c) $\epsilon = 3$

Differential privacy for small-plaintext-space data aggregation



(a) $\epsilon = 1$

(b) $\epsilon = 2$

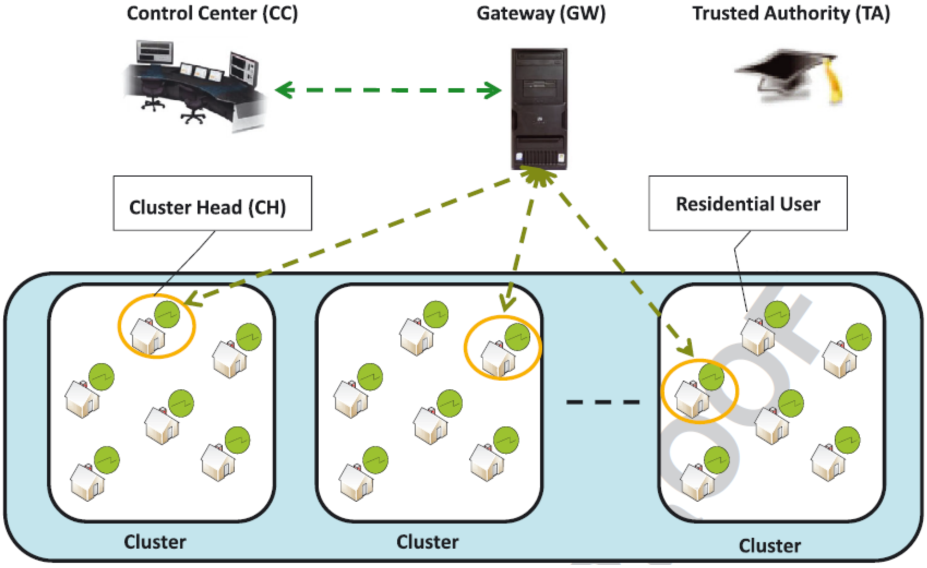
(c) $\epsilon = 3$

Differential privacy for large-plaintext-space data aggregation

(6) A Lightweight Privacy-Preserving Scheme with Data Integrity for Smart Grid Communications

Abstract

Smart grid, deemed as the next generation of power grid, can efficiently monitor, control, and predicate energy generation and consumption. However, the frequent collection of users' consumption information in smart grid may reveal user's privacy, and the tampering of smart grid communication may also impair the data integrity, subsequently affecting the precise monitoring and controlling at the control center. In this paper, to address the aforementioned challenges, we propose a lightweight data report scheme for smart grid communications, which can achieve privacy preservation and data integrity simultaneously. Specifically, an efficient pseudonym identity-based privacy-preserving report approach is proposed for the control center to obtain the fine-grained usage data of all the users while protecting user's privacy. An online/off-line hash tree-based mechanism is also designed to check and assure data integrity of communications. Because of the shifting of most time-consuming computations to off-line phase, the online process is very fast and efficient by performing merely the lightweight bottom-up hash tree verifications to check all users' data integrity concurrently. Furthermore, a topology-independent data report architecture is also structured, which is adaptable for dynamic residential users to spontaneously form clusters and efficiently report data in flocks. Extensive performance evaluation demonstrates that the proposed scheme can achieve less communication overhead and dramatically reduce computational cost in comparison with the existing schemes.



System model under consideration

Major Features and Contributions

- A lightweight pseudonym identity-based privacy-preserving data report approach is proposed. Different from the existing data aggregation schemes, in which just the sum usage data can be obtained by the control center (CC), the fine-grained usage data of all users can be obtained by CC in privacy-preserving way. Thus, provided that user's privacy is not revealed, with the detailed information, the whole system can be monitored and controlled more efficiently by CC.
- An online/off-line hash tree-based authentication and data integrity verification mechanism are designed. Most of the computations of the smart meter with limited resources could be pre-processed in off-line phase. Furthermore, source authentication and data integrity of all the received usage reports can be checked simultaneously by performing the bottom-up hash tree verification procedures.

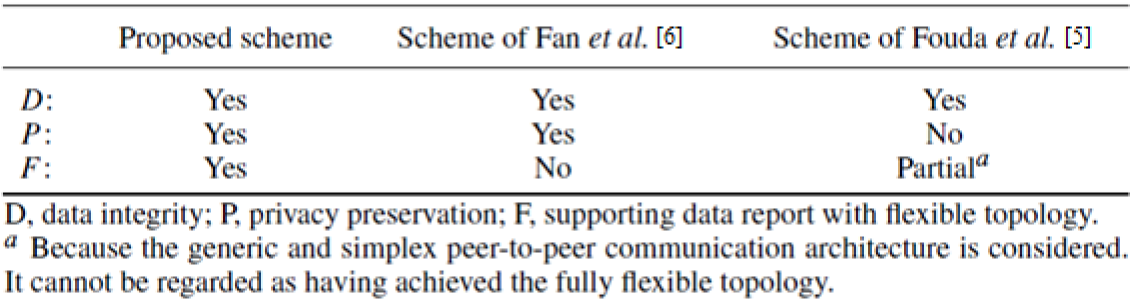
- A distributed and autonomous data collection architecture is structured. The users in the neighboring areas can form the cluster dynamically and flexibly, which makes the data report to be topology independent. And extensive performance evaluation demonstrates that the proposed architecture can achieve less communication overhead and dramatically reduce computation cost compared with the existing similar schemes.

Performance Evaluation

The proposed scheme achieves privacy preservation and data integrity simultaneously for secure data report with flexible topology of users in RA for smart grid communications. We mainly compare the performance of our proposed scheme with the state-of-the-art similar schemes [5, 6].

A. Computation Cost

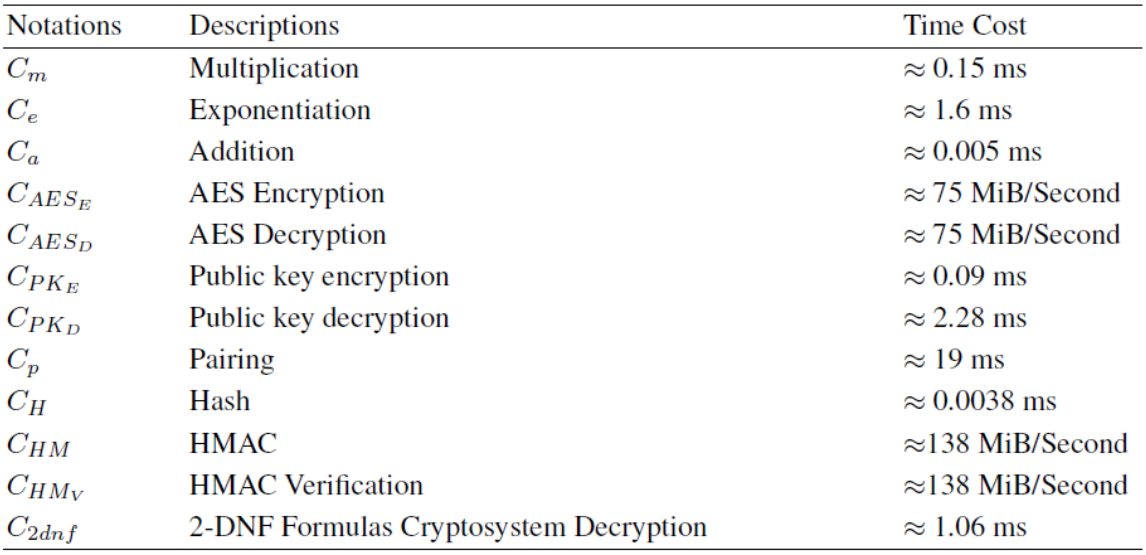
The features comparisons, time cost of operations, computation cost comparisons, and performance comparisons of computation cost are illustrated as follows.



	Proposed scheme	Scheme of Fan *et al.* [6]	Scheme of Fouda *et al.* [5]
D:	Yes	Yes	Yes
P:	Yes	Yes	No
F:	Yes	No	Partial[a]

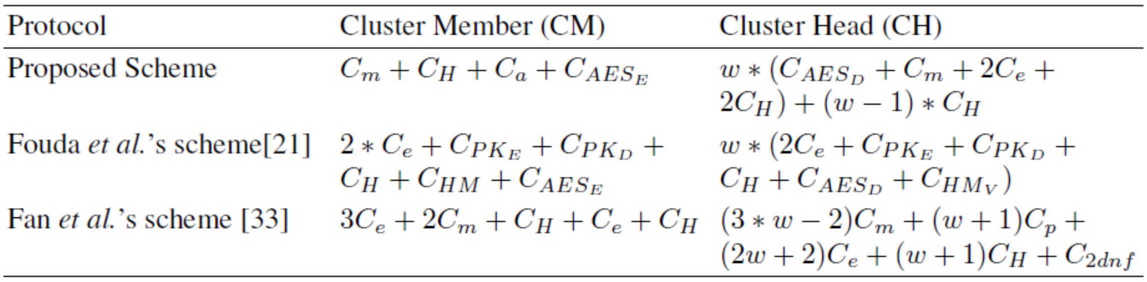
D, data integrity; P, privacy preservation; F, supporting data report with flexible topology.
[a] Because the generic and simplex peer-to-peer communication architecture is considered. It cannot be regarded as having achieved the fully flexible topology.

Feature comparison



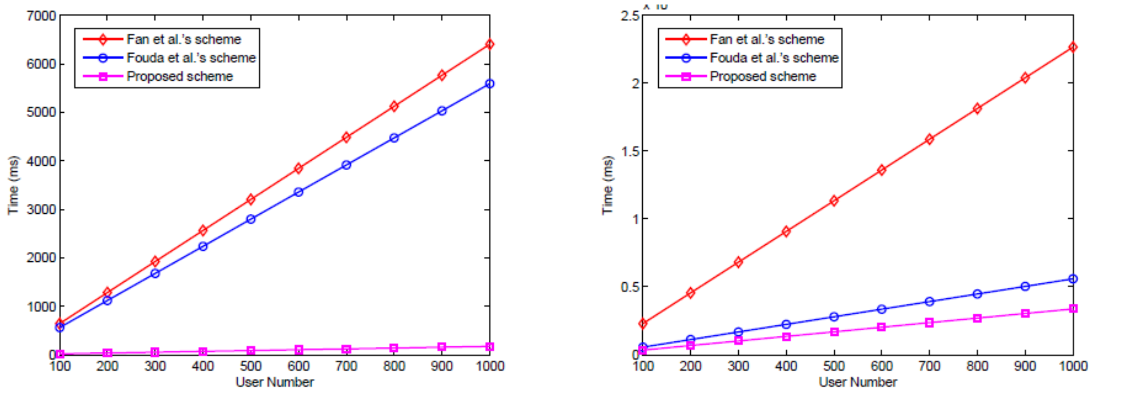
Notations	Descriptions	Time Cost
C_m	Multiplication	$\approx$ 0.15 ms
C_e	Exponentiation	$\approx$ 1.6 ms
C_a	Addition	$\approx$ 0.005 ms
C_{AES_E}	AES Encryption	$\approx$ 75 MiB/Second
C_{AES_D}	AES Decryption	$\approx$ 75 MiB/Second
C_{PK_E}	Public key encryption	$\approx$ 0.09 ms
C_{PK_D}	Public key decryption	$\approx$ 2.28 ms
C_p	Pairing	$\approx$ 19 ms
C_H	Hash	$\approx$ 0.0038 ms
C_{HM}	HMAC	$\approx$138 MiB/Second
C_{HM_V}	HMAC Verification	$\approx$138 MiB/Second
C_{2dnf}	2-DNF Formulas Cryptosystem Decryption	$\approx$ 1.06 ms

Time cost of operations



Protocol	Cluster Member (CM)	Cluster Head (CH)
Proposed Scheme	$C_m + C_H + C_a + C_{AES_E}$	$w * (C_{AES_D} + C_m + 2C_e + 2C_H) + (w-1) * C_H$
Fouda *et al.*'s scheme[21]	$2 * C_e + C_{PK_E} + C_{PK_D} + C_H + C_{HM} + C_{AES_E}$	$w * (2C_e + C_{PK_E} + C_{PK_D} + C_H + C_{AES_D} + C_{HM_V})$
Fan *et al.*'s scheme [33]	$3C_e + 2C_m + C_H + C_e + C_H$	$(3 * w - 2)C_m + (w+1)C_p + (2w+2)C_e + (w+1)C_H + C_{2dnf}$

Computation cost comparisons

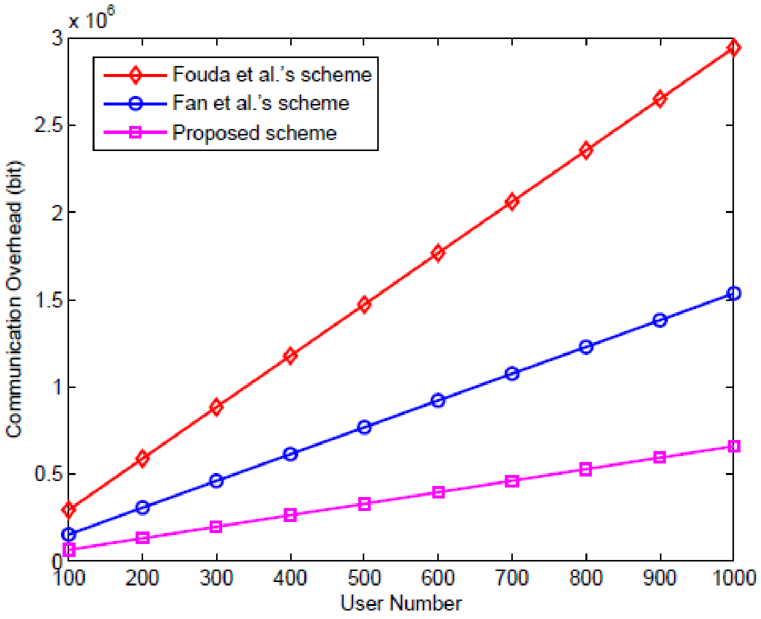


(a) Performance comparison of computation cost at cluster member (CM) side

(b) Performance comparison of computation cost at cluster head (aggregator) side

Performance comparison of computation cost

B. Communication Overhead



Performance comparison of communication overhead

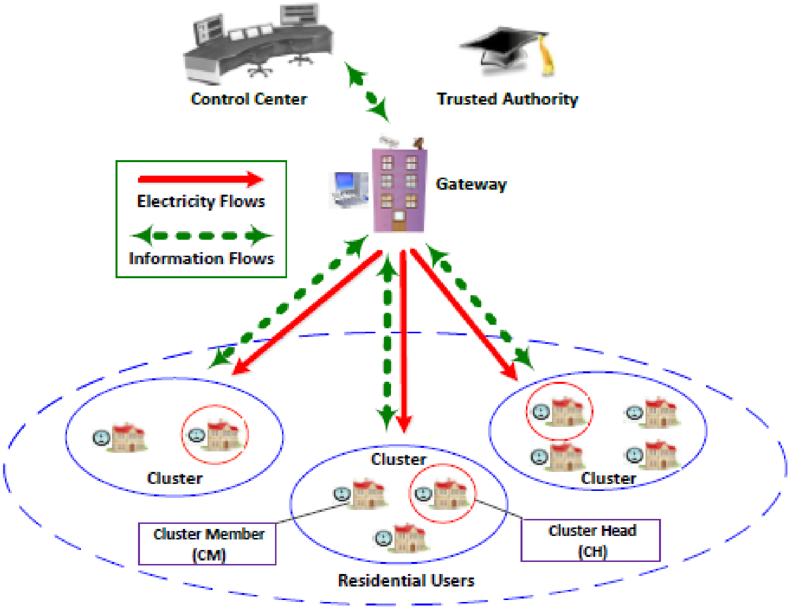
Performance evaluations show that our proposed scheme is indeed efficient in terms of computation and communication cost, which is suitable for the real-time high-frequency data report in smart grid communications.

(7) A Lightweight Data Aggregation Scheme Achieving Privacy Preservation and Data Integrity with Differential Privacy and Fault Tolerance

Abstract

To design an efficient and secure data aggregation scheme fitting real applications has been pursued by research communities for a long time. In this paper, we propose a novel secure data aggregation scheme to simultaneously achieve privacy preservation and data integrity with differential privacy and fault tolerance. Specifically, by introducing some auxiliary ciphertext subtly, a novel distributed solution for fault tolerant data aggregation is put forward to be able to aggregate the functioning smart meter measurements flexibly and efficiently for any rational number of malfunctioning smart meters with discretional long failure period. The proposed scheme also achieves a good trade-off of accuracy and security of differential privacy for arbitrary number of malfunctioning smart meters. In the proposed scheme, a novel efficient authentication mechanism is also proposed to generate and share session keys in a non-interactive way, which is leveraged for AES encryption to achieve source authentication and data integrity of the transmitted data. Furthermore, through decentralizing the

computational overhead and the authority of the hub-like entity of the gateway, the security of our proposed scheme is enhanced and the efficiency is improved significantly. Finally, extensive performance evaluations are conducted to illustrate that the proposed data aggregation scheme outperforms the state-of-the-art similar schemes in terms of computation complexity, communication cost, robustness of fault tolerance, and utility of differential privacy.



System model under consideration

Major Features and Contributions

- By introducing auxiliary ciphertext subtly, we put forward a novel distributed solution for fault tolerant data aggregation. Unlike most of the existing similar works, which depend on the central trust authority to trace and separate the malfunctioning smart meters from the functioning ones to be able to aggregate the smart meter measurements in case of report failures, our proposed scheme supports fault tolerance of malfunctioning smart meters without the participation and restriction of any external factors. Specifically, utilizing the auxiliary ciphertexts, CC can obtain the aggregation of the functioning smart meters flexibly and efficiently for any rational number of malfunctioning smart meters with arbitrary long failure.
- Observing the fact that user's private data may often suffer from differential attacks, our proposed scheme provides differential privacy by adding appropriate noises chosen from Symmetric Geometric distribution to the aggregation data by GW. To the best of our knowledge, most of the existing similar works cannot support differential privacy and fault tolerance at the same time. A handful of literatures trying to address this problem only consider the scenarios that there is small amount (or fixed maximum number) of malfunctioning smart meters to be able to add appropriate noises to support differential privacy. Our scheme supports differential privacy and fault tolerance simultaneously, and achieves a good trade-off of accuracy and security of differential privacy for arbitrary number.
- By integrating a pair of identities and private/public keys of two communication parties, and current time slot for data report, a novel efficient authentication technique is proposed to flexibly generate and share session keys in non-interactive way. The shared session key is leveraged for AES encryption to achieve source authentication and data integrity of transmitted data. The security analysis and performance evaluation indicate that the proposed mechanism can efficiently and

effectively prevent the malicious adversary from impairing and polluting (e.g., modify, forge, inject, reply and/or delay, etc.) the transmitted data.

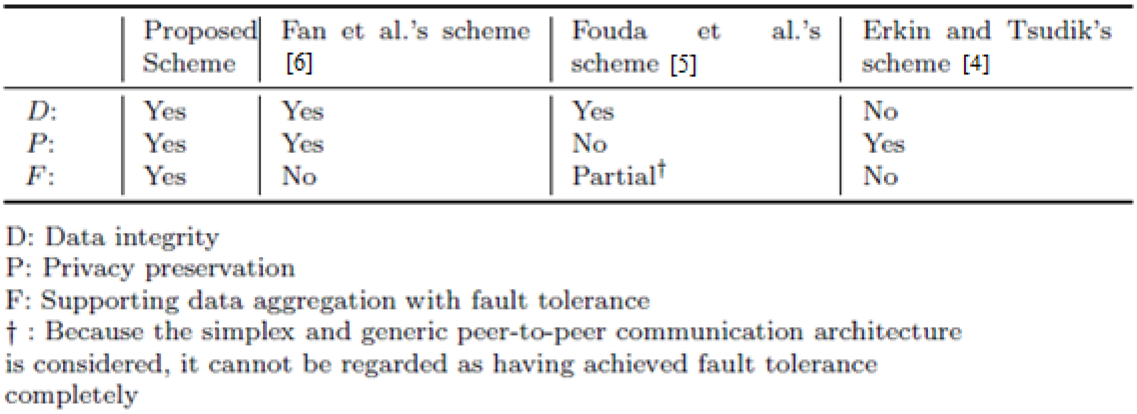
- Through decentralizing the computational overhead and the power of the hub-like entity GW, which is usually with limited computation resources and is semi-trust, the security of our proposed scheme is enhanced and the efficiency is improved significantly. Specifically, only the encryption of the usage data and the auxiliary ciphertext are aggregated and processed beforehand by at least two users, respectively, can they be reported to GW. In addition, through comparative performance analysis, we demonstrate that our proposed data aggregation scheme outperforms the state-of-the-art similar schemes [3] in terms of computation complexity, communication cost, robustness of fault tolerance, and utility of differential privacy.

Performance Evaluation

The proposed scheme achieves privacy preservation and data integrity simultaneously for secure data aggregation with differential privacy and fault tolerance for smart grid communications. We mainly compare the performance of our proposed scheme with the state-of-the-art similar schemes [4, 5, 6].

A. Computation Complexity

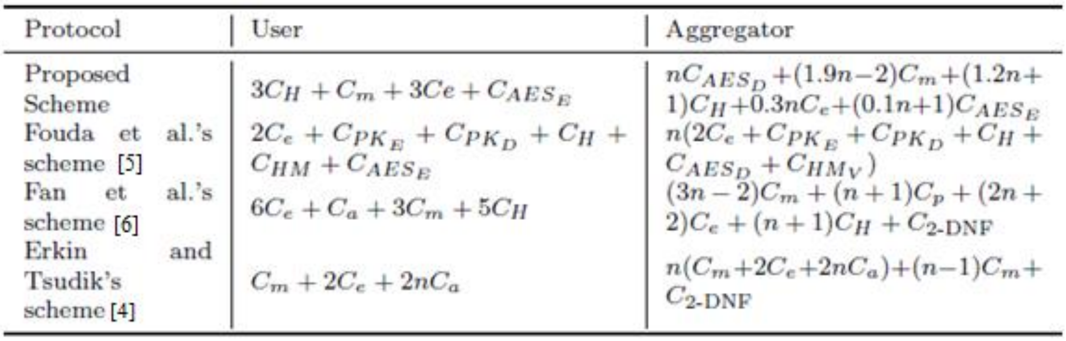
The features comparisons, computation cost comparisons, and performance comparisons of computation cost are as follows.



	Proposed Scheme	Fan et al.'s scheme [6]	Fouda et al.'s scheme [5]	Erkin and Tsudik's scheme [4]
D:	Yes	Yes	Yes	No
P:	Yes	Yes	No	Yes
F:	Yes	No	Partial†	No

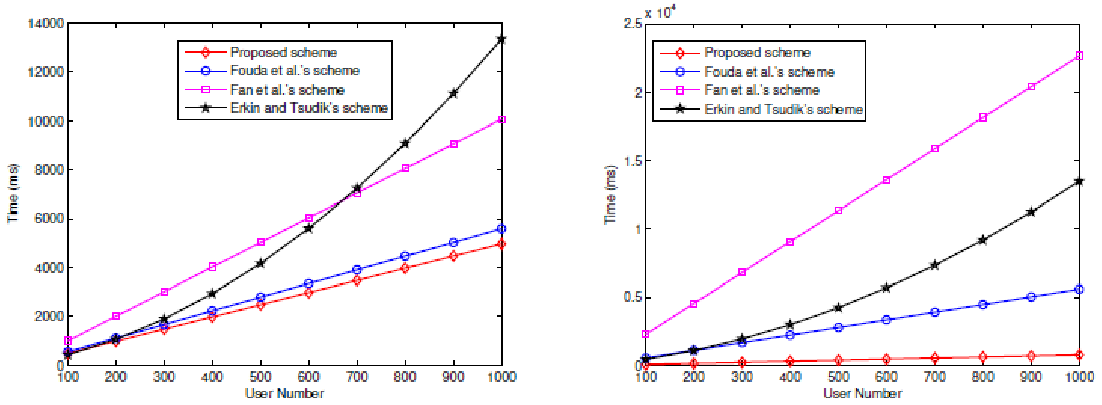
D: Data integrity
P: Privacy preservation
F: Supporting data aggregation with fault tolerance
† : Because the simplex and generic peer-to-peer communication architecture is considered, it cannot be regarded as having achieved fault tolerance completely

Feature comparison



Protocol	User	Aggregator
Proposed Scheme	$3C_H + C_m + 3Ce + C_{AES_E}$	$nC_{AES_D} + (1.9n-2)C_m + (1.2n+1)C_H + 0.3nC_e + (0.1n+1)C_{AES_E}$
Fouda et al.'s scheme [5]	$2C_e + C_{PK_E} + C_{PK_D} + C_H + C_{HM} + C_{AES_E}$	$n(2C_e + C_{PK_E} + C_{PK_D} + C_H + C_{AES_D} + C_{HM_V})$
Fan et al.'s scheme [6]	$6C_e + C_a + 3C_m + 5C_H$	$(3n-2)C_m + (n+1)C_p + (2n+2)C_e + (n+1)C_H + C_{2\text{-DNF}}$
Erkin and Tsudik's scheme [4]	$C_m + 2C_e + 2nC_a$	$n(C_m + 2C_e + 2nC_a) + (n-1)C_m + C_{2\text{-DNF}}$

Computation cost comparisons

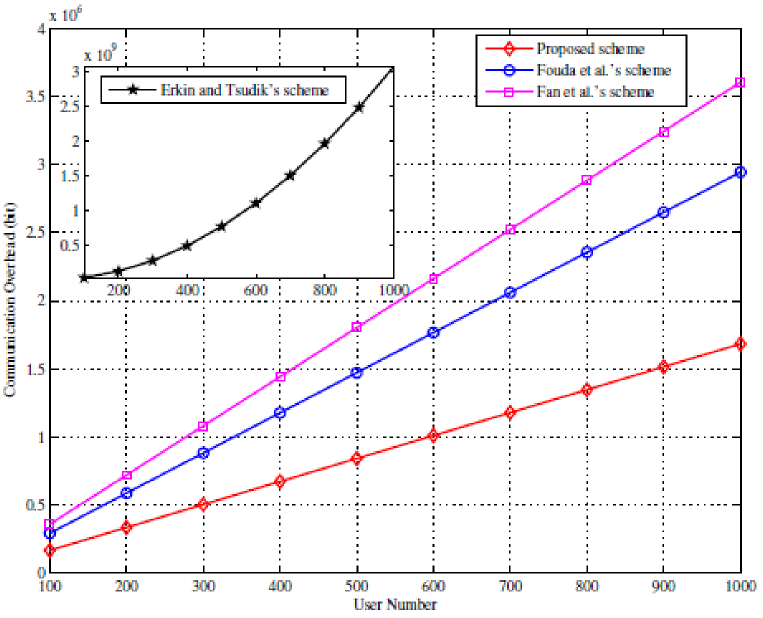


(a) Performance comparison of computation cost at user side

(b) Performance comparison of computation cost at aggregator side

Performance comparison of computation cost

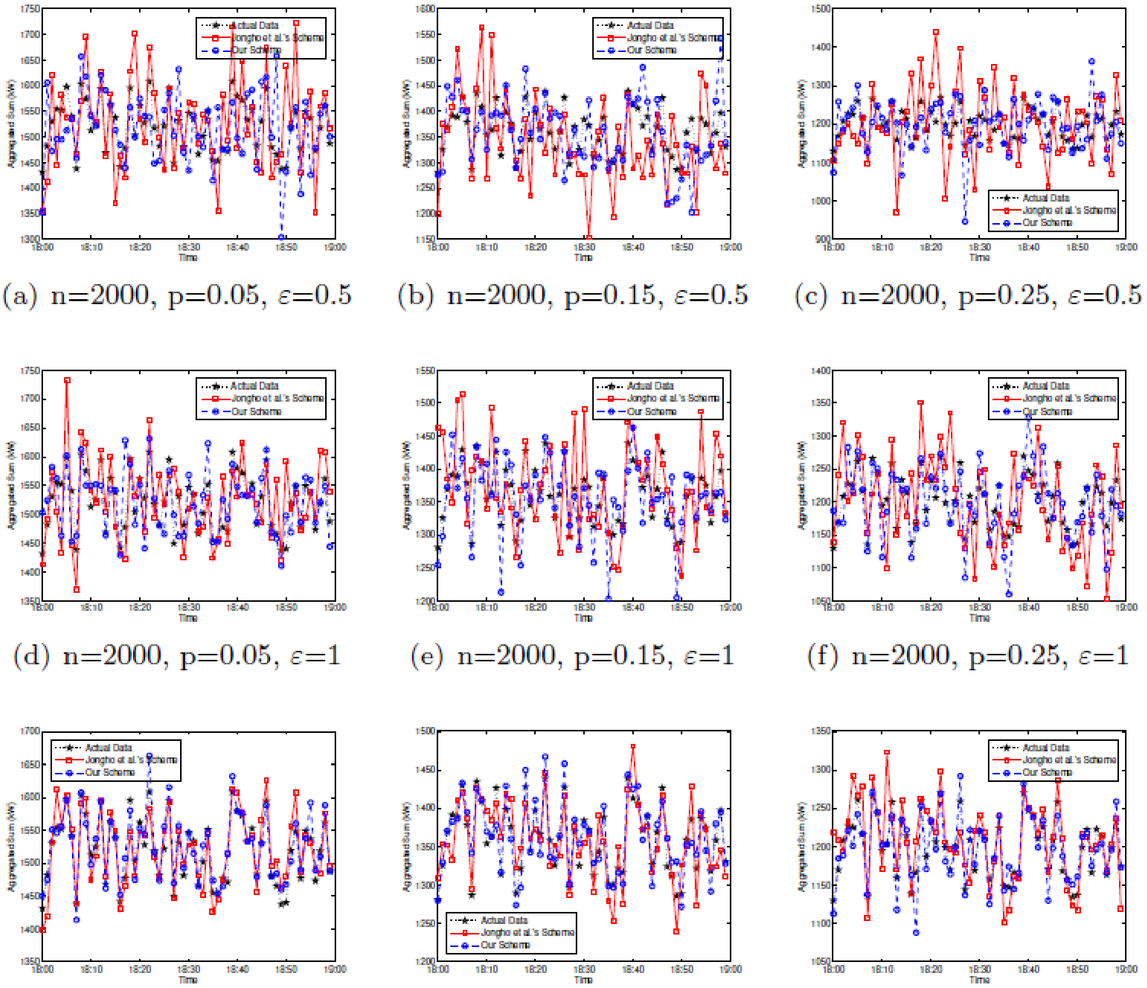
B. Communication Cost



Performance comparison of communication overhead

C. Utility of Differential Privacy

Based on real data of 2000 households, we compare the utility of differential privacy of our proposed scheme with the state-of-the-art one [3]. The following figures illustrate the traces of the actual total measurements, the noisy counterparts of both [3] and our proposed scheme, for the different parameters, where in each of the figure, n and p denote the total number of the household, and the different ratio of malfunctioning smart meters, respectively. As it can be seen from the figures, the larger the number of p, the more accurate of our scheme comparing with the scheme of [3].



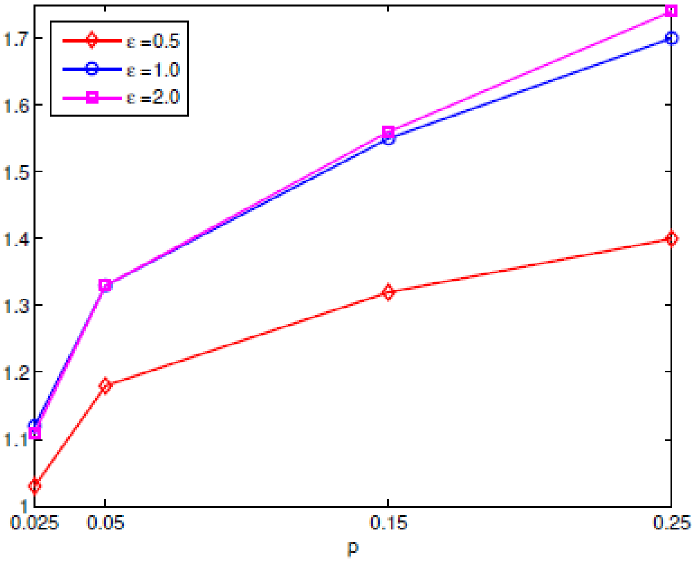
(a) n=2000, p=0.05, ε=0.5 (b) n=2000, p=0.15, ε=0.5 (c) n=2000, p=0.25, ε=0.5

(d) n=2000, p=0.05, ε=1 (e) n=2000, p=0.15, ε=1 (f) n=2000, p=0.25, ε=1

(g) n=2000, p=0.05, ε=2 (h) n=2000, p=0.15, ε=2 (i) n=2000, p=0.25, ε=2

Comparison of noisy total consumption between the proposed aggregation protocol and Jongho et al.'s protocol [3]

Let the 1-h root-mean-square-error (RMSE) of Jongho et al.'s protocol [3] and our proposed scheme be γ_1 and γ_2 respectively. The ratios of $\gamma = \frac{\gamma_1}{\gamma_2}$ with p under different privacy level ε are depicted in the following figure, which shows that comparing with [3], our proposed scheme always achieves better utility due to much lower errors in each circumstance.



Comparison of 1-h RMSE between the proposed aggregation protocol and Jongho et al.s protocol [3]

Through the above comparison, we can see that our proposed data aggregation scheme provides higher utility (i.e., low error) in terms of differential privacy than the state-of-the-art data aggregation scheme of [3].

References

[1] Garcia FD, Jacobs B. Privacy-friendly energy metering via homomorphic encryption, Proceedings of the 6th International Conference on Security and Trust Management, STM'10, Springer-Verlag: Berlin Heidelberg, 2011; 226–238.

[2] Li F, Luo B, Liu P. Secure information aggregation AQ5 for smart grids using homomorphic encryption, 2010 First IEEE International Conference on Smart Grid Communications (SmartGridComm), IEEE, 2010; 327–332.

[3] J. Won, C. Y. Ma, D. K. Yau, and N. S. Rao, "Proactive fault-tolerant aggregation protocol for privacy-assured smart metering," in INFOCOM 2014. IEEE, 2014, pp. 2804–2812.

[4] Shi, E., Chan, T. H. H., Rieffel, E. G., Chow, R., & Song, D. Privacy-Preserving Aggregation of Time-Series Data. In NDSS 2011, Vol. 2, No. 3, p. 4.

[5] Fouda MM, Fadlullah ZM, Kato N, Lu R, Shen X. A lightweight message authentication scheme for smart grid COMMUNICATIONS. IEEE Transactions on Smart Grid 2011; 2(4):675–685.

[6] Fan CI, Huang SY, Lai YL. Privacy-enhanced data aggregation scheme against internal attackers in smart grid. IEEETransactions on Industrial Informatics 2014; 10(1):666–675.

[7] R. Lu, X. Liang, X. Li, X. Lin, and X. Shen, "EPPA: an efficient and privacy-preserving aggregation scheme for secure smart grid communications," IEEE Trans. Parallel Distrib. Syst., vol. 23, no. 9, pp. 1621–1631, 2012. [Online]. Available at http://dx.doi.org/10.1109/TPDS.2012.86

List of Publications and Significant Collaborations that resulted from your AOARD supported project: In standard format showing authors, title, journal, issue, pages, and date, for each category list the following:

a) papers published in peer-reviewed ournals,
b) papers published in peer-reviewed conference proceedings,
c) papers published in non-peer-reviewed journals and conference proceedings, d) conference presentations without papers,
e) manuscripts submitted but not yet published, and
f) provide a list any interactions with industry or with Air Force Research Laboratory scientists or significant collaborations that resulted from this work.

Attachments: Publications a), b) and c) listed above if possible.

DD882: As a separate document, please complete and sign the inventions disclosure form.

Important Note: If the work has been adequately described in refereed publications, submit an abstract as described above and refer the reader to your above List of Publications for details. If a full report needs to be written, then submission of a final report that is very similar to a full length journal article will be sufficient in most cases. This document may be as long or as short as needed to give a fair account of the work performed during the period of performance. There will be variations depending on the scope of the work. As such, there is no length or formatting constraints for the final report. Keep in mind the amount of funding you received relative to the amount of effort you put into the report. For example, do not submit a $300k report for $50k worth of funding; likewise, do not submit a $50k report for $300k worth of funding.

Include as many charts and figures as required to explain the work.

List of Publications and Significant Collaborations that resulted from your AOARD supported project: In standard format showing authors, title, journal, issue, pages, and date, for each category list the following:

a) papers published in peer-reviewed journals,

[1] **Journal name:** Security and Communication Networks, Vol.8, No. 15, pp. 2494-2506, Nov 2015.
Title: "PDA: A Privacy-Preserving Dual-Functional Aggregation Scheme for Smart Grid Communications A Lightweight Data Aggregation Scheme Achieving Privacy Preservation and Data Integrity with Differential Privacy and Fault Tolerance"
Date: 23 Jan 2015
Authors: Chen Li, Rongxing Lu, Hui Li, Le Chen, and Jie Chen

[2] **Journal name:** IEEE Internet of Things, Vol.2, No. 3, pp. 248-258, Jun 2015.
Title: "A New Differentially Private Data Aggregation with Fault Tolerance for Smart Grid Communications"
Date: 13 Mar 2015
Authors: Haiyong Bao, and Rongxing Lu

[3] **Journal name:** Concurrency and Computation: Practice and Experience, DOI: 10.1002/cpe.3527.
Title: "A Lightweight Privacy-Preserving Scheme with Data Integrity for Smart Grid Communications"
Date: 28 May 2015
Authors: Haiyong Bao, and Le Chen

[4] **Journal name:** Peer-to-Peer Networking and Applications, accepted
Title: "A Lightweight Data Aggregation Scheme Achieving Privacy Preservation and Data Integrity with Differential Privacy and Fault Tolerance Privacy-Preserving Time-Series Data Aggregation for Internet of Things"
Date: to appear
Authors: Haiyong Bao, Rongxing Lu

b) papers published in peer-reviewed conference proceedings,

[1] **Conf. name:** IEEE ICC'15, DOI: 10.1109/ICC.2015.7249482
Title: "DDPFT: Secure Data Aggregation Scheme with Differential Privacy and Fault Tolerance"
Date: 8-12 Jun 2015
Authors: Haiyong Bao, and Rongxing Lu

[2] **Conf. name:** IEEE GLOBECOM'15
Title: "A Novel Privacy-Preserving Set Aggregation Scheme for Smart Grid Communications"
Date: 6-10 Dec 2015
Authors: Rongxing Lu, Khalid Alharbi, Xiaodong Lin, and Cheng Huang

c) papers published in non-peer-reviewed journals and conference proceedings,

None

d) conference presentations without paper

None

e) manuscripts submitted but not yet published, and

[1] **Book name:** Security and Privacy in Internet of Things (IoTs): Models, Algorithms, and Implementations, accepted

Title: "Privacy-Preserving Time-Series Data Aggregation for Internet of Things"
Date received:
Authors: Rongxing Lu, Xiaodong Lin, Cheng Huang, Haiyong Bao

f) provide a list any interactions with industry or with Air Force Research Laboratory scientists or significant collaborations that resulted from this work.

None

4. Invited talks (event name, title, date):

None

5. Award for best paper, best poster (title, date):

- ITS Student Best Paper Award, Intelligent Transport Systems (ITS) Summit Singapore 2015, Singapore, 2015
- "Energy-Theft Detection Issues for Advanced Metering Infrastructure in Smart Grid", TSINGHUA Science and Technology, Vol.19, No. 2, pp. 105-120, 2014. (received the 2014 Best Paper Award of TSINGHUA Science and Technology)
- "An Efficient Data-driven Particle PHD Filter for Multitarget Tracking", IEEE Transactions on Industrial Informatics, Vol.9, No. 4, pp. 2318-2326, 2013. (received the 2014 IEEE IES Student Best Paper Award and the award ceremony was held during the society's flagship conference IECON 2014 in Dallas, Texas, USA, Oct. 29 - Nov.1, 2014)

6. Award of fund received related to your research efforts (name, amount, date):

- 2015-2018 Project-PI, Economic Development Board (EDB): Development of NTU/ NXP- Intelligent Transport Systems Test-Bed (V2X-RP4: Security enhanced technologies for IEEE 802.11p), S$ 550,000
- 2015-2016 PI, CoE Proposal Preparatory Grant (College of Engineering): Secure Fog Computing System Design and Implementation, S$ 50,000
- 2015-2017 PI, MOE AcRF Tier 1: Toward Secure and Privacy-Preserving Computing in Mobile Cloud, S$ 99,803